



Published in final edited form as:

*Int J Audiol.* 2015 February ; 54(0 1): S19–S29. doi:10.3109/14992027.2014.979300.

## STIMULUS AND TRANSDUCER EFFECTS ON THRESHOLD

Gregory A. Flamme<sup>a</sup>, Kyle Geda<sup>a</sup>, Kara McGregor<sup>a</sup>, Krista Wyllys<sup>a</sup>, Kristy K. Deiters<sup>a</sup>, William J. Murphy<sup>b</sup>, and Mark R. Stephenson<sup>b</sup>

<sup>a</sup>Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, MI, USA

<sup>b</sup>Division of Applied Research and Technology, National Institute for Occupational Safety and Health, Cincinnati, Ohio, USA

### Abstract

**Objective**—This study examined differences in thresholds obtained under Sennheiser HDA200 circumaural earphones using pure tone, equivalent rectangular noise bands, and 1/3 octave noise bands relative to thresholds obtained using Telephonics TDH-39P supra-aural earphones.

**Design**—Thresholds were obtained via each transducer and stimulus condition six times within a 10-day period.

**Study Sample**—Forty-nine adults were selected from a prior study to represent low, moderate, and high threshold reliability.

**Results**—The results suggested that (1) only small adjustments were needed to reach equivalent TDH-39P thresholds, (2) pure-tone thresholds obtained with HDA200 circumaural earphones had reliability equal to or better than those obtained using TDH-39P earphones, (3) the reliability of noise-band thresholds improved with broader stimulus bandwidth and was either equal to or better than pure-tone thresholds, and (4) frequency-specificity declined with stimulus bandwidths greater than one Equivalent Rectangular Band, which could complicate early detection of hearing changes that occur within a narrow frequency range.

**Conclusions**—These data suggest that circumaural earphones such as the HDA200 headphones provide better reliability for audiometric testing as compared to the TDH-39P earphones. These data support the use of noise bands, preferably ERB noises, as stimuli for audiometric monitoring.

### Keywords

Audiometry; noise-induced hearing loss; reliability; occupational health

---

Correspondence to: Gregory A. Flamme, Department of Speech Pathology and Audiology, Western Michigan University, 1903 W. Michigan Ave., Kalamazoo MI, 49008, greg.flamme@wmich.edu.

#### Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of Centers for Disease Control and Prevention (CDC) or the National Institute for Occupational Safety and Health (NIOSH). Mention of any company or product does not constitute endorsement by CDC or NIOSH.

## Background

Threshold audiometry was one of the first methods established for the measurement of hearing sensitivity (e.g., Fletcher & Wegel, 1922), and it remains the gold standard procedure (Engdahl et al., 2012). The threshold audiogram provides frequency-specific comparisons of a listener's response against the responses that would typically be expected from a young population with normal hearing (i.e., 0 dB Hearing Level, or HL).

Band-limited stimuli are necessary to provide frequency-specific information, and pure tones were initially adopted because they represent the minimum possible bandwidth and are easy to generate. The reliability of pure tone thresholds obtained using the TDH-39P supra-aural earphone is poorer in the high frequencies (Flamme et al., 2014) where stimulus wavelengths are comparable to the distance from the transducer diaphragm to the eardrum and standing waves are possible. In addition, ringing in the ears (i.e., tinnitus) tends to have a tone-like quality that can be confused with the tone, which complicates the interpretation of pure tone test results for listeners with tinnitus.

Regular audiometric monitoring is a key component of hearing conservation programs. The purpose of monitoring audiometry is to identify changes from baseline threshold and quickly determine whether the change is associated with excess exposure to noise or other ototoxicants before any change in hearing interferes with performance in daily life. High reliability, therefore, is crucial to the task of identifying changes in hearing sensitivity as early as possible. The reliability of pure tone threshold audiometry with TDH-39P earphones is moderately good, but improvements in reliability in the high frequencies are desired (Flamme et al., 2014).

The attenuation of hearing protectors is conventionally measured using the differences between thresholds with and without the protection device in place. High reliability of the measurement is also important for hearing protector measurements. Narrow bands of noise have long been used as the preferred stimuli for assessment of hearing protector attenuation, partly due to the common need to test hearing protectors in sound fields where a uniform sound field for tonal stimuli would be nearly impossible.

Lab measurements of earplug attenuation tend to overestimate the amount of attenuation observed among workers in practice (Berger et al., 1998). This could be partly due to differences in training or motivation and partly due to the application of research procedures that cannot be duplicated in the field. However, field-based systems for checking the attenuation of hearing protectors have been developed (Murphy, 2013) using high-quality low-cost audio systems (e.g., laptop sound cards, tablet computers), and the technical requirements of those systems can be similar to the technical requirements for threshold audiometry. It is possible to devise a field-based system that combines audiometric monitoring and individualized assessment of earplug attenuation into a low-cost and efficient procedure. It would be necessary in such a system to use instrumentation and procedures capable of assessing occluded and unoccluded thresholds. At minimum, this requirement implies the use of circumaural earphones for assessing the attenuation of

earplugs. Further, this system should also produce results comparable to conventional pure tone thresholds and measures of hearing protector attenuation.

The increased bandwidth of noise stimuli could reduce the extent to which narrow frequency regions of reduced audibility are observed and this could also lead to underestimates (i.e., artificial improvements) in threshold on steeply-sloped segments of the audiogram. The pure tone stimulus will excite primarily the portion of the basilar membrane surrounding the pure tone frequency. However, the auditory filter is spread continuously about the point that serves this frequency. The Equivalent Rectangular Band (ERB, Glasberg & Moore, 1990) is intended to represent a rectangular filter shape that has the same area as the auditory filter. Auditory filter measurements suggest that the shape of the auditory filter follows a rounded exponential or compressive gammachirp curve (Unoki et al., 2006). Given that a stimulus with a rectangular spectrum is used to approximate a non-rectangular auditory filter shape, it is possible that neighboring auditory filters could be excited by rectangular bands, and responses from adjacent auditory filters could lead to apparent improvements in sensitivity if better sensitivity is present in the adjacent filters. This would result in an apparent “filling” of audiometric notches, which would be exhibited by reduced absolute slope between neighboring frequencies. One-third octave band (1/3 OB) signals have also been used to obtain frequency-specific threshold information (e.g., Cox & McDaniel, 1986). One could expect that 1/3 OB signals would reduce slopes between neighboring audiometric frequencies more than ERB signals because the ERBs are narrower than 1/3 OB signals.

This study had three objectives. The first objective was to determine whether thresholds obtained on the HDA200 earphones using noise bands are exchangeable, or can be transformed into, equivalent pure tone thresholds obtained with TDH-39P supra-aural audiometric earphones. The second objective was to determine whether thresholds obtained with the HDA200 earphone results in substantially different reliability than pure tone thresholds obtained with conventional supra-aural audiometric earphones. Finally, we conducted an exploratory assessment of whether the use of noise bands influenced the slope of the audiogram in cases of large threshold changes between neighboring frequencies. These objectives contribute to a long-term goal of evaluating the feasibility of conducting audiometric monitoring and field testing of earplug attenuation using a single stimulus and earphone model.

## Method

### Participants

The participants in this study were a subset of 49 participants who previously completed a larger study of the reliability of pure tone thresholds (Flamme, et al. 2014). Participants who completed the prior study were divided into three reliability groups (high, medium, low) of equal size based on mean squared deviation to the mean threshold (across ears, stimulus frequencies, and a total of ten separate tests). Participants were selected on the basis of known threshold reliability in order to ensure generalizability of results to the population. Invitations to participate in the current study were issued to obtain approximately equal numbers of men and women in each reliability category, which led to a study sample of 26 men and 23 women. Participants were between 20 and 69 years of age, and the majority (60

%) of the sample was between 40 and 59 years old. A total of 11 participants were between the ages of 20 and 39, and eight participants were between the ages of 60 and 69 years of age. No systematic relationship between decade of age and reliability category was observed in this sample (Fisher's exact  $p = 0.738$ ). One ear was selected at random for testing in the current study.

## Stimuli

Pure tones, one-third octave noise bands, and noises of one equivalent rectangular bandwidth (ERB) were used in this study. Pure tones were generated using Matlab (Mathworks, Inc., Natick, MA). Noise bands were generated by first producing a 120-second Gaussian noise to produce a signal with a uniform spectrum density. The random noise was then filtered digitally (using a 100,000-order finite impulse response filter) to produce signals with very steep rejection slopes (e.g., 1500 dB/octave) and high stopband attenuation in order to maximize the extent to which participant responses represented performance at a restricted frequency region. The ERB bandwidths were obtained using the equation derived by Glasberg & Moore (1990) for moderate sound levels:

$$ERB = 24.7(4.37F + 1) \quad \text{Equation 1}$$

where *ERB* is the equivalent rectangular bandwidth, in Hz, surrounding the frequency *F*, in kHz. The noise bands were logarithmically centered on the nominal stimulus frequency.

The narrow band noise spectra had flat passbands initially. However, the spectrum of the electronic stimuli would be filtered by the frequency response of the HDA200 earphone (see Figure 1) and external ear, so the stimuli were filtered to match the inverse of the HDA200 frequency response (averaged across the right and left transducer). Note that although there were some differences across transducers within the pair, the frequency response shape was similar and we judged that a single transfer function would be sufficient. The HDA200 frequency response was measured at the output of an IEC-60711 ear simulator mounted in KEMAR, so the passbands of the filtered noises were designed to be flat at the level of the average human adult eardrum. Finally, the stimuli were re-scaled for equal root-mean-square amplitudes and to conform to the .wav sound file format with 16 bit resolution, a 44.1 kHz sampling rate, and maximum absolute values less than 1.0. The spectra of the noise band stimuli are represented in Figure 2.

Pure tone stimuli were delivered to both the TDH-39P and HDA200 earphones, and noise bands were delivered only to the HDA200 earphones. The TDH-39P earphones were not used for noise band testing because these earphones have erratic frequency responses above their resonance frequency (just below 6 kHz). In addition, the supra-aural design of the TDH-39P is incompatible with field testing of earplug attenuation.

## Instrumentation

The Nelson Acoustics Audiometric Research Tool (ART) software program (VIAcoustics, Inc., Austin, Texas) was used for threshold tests. This was chosen because it provided a single well-understood platform for testing thresholds via multiple stimuli and transducers and because it provided access to the presentation and response history associated with each

observed threshold. The ART software was run using a National Instruments (NI) embedded controller system (PXIe-8133) mounted within an NI PXIe chassis. The NI PXI-4461 dynamic signal analyzer module was used for digital-analog conversion. Signals were then routed via a switchbox to either Telephonics TDH-39P or Sennheiser HDA200 earphones. Specific ART configuration files were used to route the signal into the appropriate (left or right) channel, identify stimulus .wav files, presentation parameters (200 msec on-time, 25 ms linear ramp, 50 % duty cycle), and load the necessary calibration offsets for the combination of earphone, channel and stimulus. Participants used a hand-held pushbutton to respond, and pushbutton status was monitored using a VIAcoustics REATmaster response switch interface and an NI PXI-6221 data acquisition module within the chassis.

Routine calibration was accomplished using a GRAS Type 43AA test fixture (GRAS Sound and Vibration, Holte, Denmark), which was outfitted with a GRAS IEC-318 ear simulator (Model RA0039). The ear simulator microphone output was conditioned using a GRAS Type 26AC preamplifier and routed to a Larson-Davis System 824 sound level meter (Larson Davis, Inc., Provo, Utah). Calibration checks with HDA200 earphones were conducted using a flat plate adapter, and calibration checks with TDH-39P earphones were conducted with the MX41A/R cushion coupled to the plastic ring of the ear simulator and the flat plate removed. Alignment marks were attached to the flat plate to facilitate consistent placement of the HDA200. High-tension springs were mounted on the Type 43AA clamp arm to ensure adequate (900 g) coupling force to the test fixture. All threshold tests were conducted in a double-walled sound booth meeting ANSI S3.1 (1999) ambient noise specifications for testing with ears uncovered.

## Procedure

**Calibration**—All stimuli were calibrated using the reference equivalent threshold SPL (RETSPL) values provided in ANSI S3.6 (2010). The noise band and pure tone stimuli presented via the HDA200 earphones were presented at an equivalent overall level. During the data collection period, overall levels for pure tones were checked twice daily, before and after testing. Across daily calibration measurements ( $n=146$ ), mean levels in the ear simulator matched corresponding RETSPL targets within 0.2 dB and 0.4 dB for the TDH-39P and HDA200 earphones, respectively. Observed levels during daily calibration measurements were more variable for the HDA200 earphones than the TDH-39P, particularly at 3 and 4 kHz (Table 1). No changes to calibration offsets were made during the data collection period

**Data collection sessions**—The ART software followed a modified Hughson-Westlake protocol, wherein threshold was specified as the lowest level at which responses were obtained to 50 % or more presentations with a minimum of three ascending trials. The threshold search phase began at 30 dB HL and descended by 10 dB in cases if a listener response was obtained, or increased by 20 dB if no response was obtained at the initial level. Upon completion of the search phase (i.e., once the participant's response suggested a change in stimulus audibility), a 5-dB ascending step and a 10-dB descending step was used. The ART software was configured to present a maximum of four tone pulses and listeners were expected to respond within a 1.5 second response window, which began 300 msec after

the first pulse onset. A random (uniform distribution) delay of 0.2 to 1 second was inserted between presentations to reduce the predictability of stimulus onset.

Data were collected over four data sessions per participant. The first session included a description of the study, documentation of informed consent, completion of history and demographic questionnaires, scheduling future appointments, and bilateral video-otoscopy. The remaining sessions were the same as one another with the exception that the sequences of the threshold tests were randomized to avoid order effects. In addition, the sessions were the same as used in the 8 kHz test-retest reliability study (see Flamme et al., 2014), with the exceptions that only one ear was tested and more audiograms were obtained per visit. After the first session, participants completed a daily questionnaire and conventional otoscopy was performed to rule out changes to the ear canal, cerumen or middle ear status. Then conventional 0.226 kHz tympanometry, wideband absorbance, and wideband tympanograms were obtained twice bilaterally. These procedures all took place in a quiet room, but not in a sound booth. The participant was then asked to enter the sound booth, instructions were given, the appropriate earphones were placed over the participant's ears according to a randomization schedule and the audiogram was obtained. All thresholds were obtained automatically using ART. Following each audiogram, the earphones were removed from the participant's ears by one of the investigators, and the participant was given a one- to two-minute break before the next test was conducted.

The four test conditions included in this study (i.e., TDH-39P tones, HDA200 tones, HDA200 ERB noises, HDA200 1/3-octave noises) were presented in random order within each trial. The random order was selected via a random permutation. A new random permutation was drawn for each of the 24 tests completed per participant completing the protocol.

## Data Analyses

In addition to general descriptive analyses, the data from this study were analyzed using models that accounted for the correlated nature of the data. For example, thresholds were obtained twice per visit (i.e., tests nested within visit), and each participant completed three visits (i.e., visits nested within participants). Observations obtained during the same visit were potentially more strongly related to one another than either will be to tests obtained during different visits. Observations obtained from one participant were also considered likely to be more strongly related to each other than they will be to tests obtained from different participants. This correlation structure was included in the analyses using multilevel models, where observations (level 1) were nested within tests (level 2), which were nested within visits (level 3), and those were nested within participants (level 4).

Multilevel models, which are also known as mixed models, are linear models with the general form:

$$y = \mu + X\beta + Zu + \varepsilon \quad \text{Equation 2}$$

where  $y$  represents the vector of responses on the dependent variable,  $\mu$  represents a constant (intercept),  $X$  represents a matrix of fixed independent variable values,  $\beta$  represents the

vector of regression coefficients for the fixed independent variables,  $\mathbf{Z}$  represents a matrix of random factors such as participant, visit and test,  $\mathbf{u}$  represents the vector of regression coefficients for the random factors, and  $\epsilon$  represents residual error. The structure of Equation 2 is given for general linear models and this structure has been applied to logistic and other generalized linear models (Rabe-Hesketh & Skrondal, 2012). Multilevel models allow for the assessment of fixed factors while controlling for the influence of random factors.

Stata v. 12 software (StataCorp, College Station, Texas) was used for multilevel data analyses. Although thresholds obtained using a 5-dB step are ordinal categorical variables, we analyzed thresholds as if they were continuous variables because the observed range of threshold values was comparatively large and because thresholds represent an underlying continuum. In order to overcome the violation of the assumption of a continuous dependent variable, robust (sandwich-based) standard errors (Huber, 1967) were used in multilevel analyses treating threshold as a continuous variable. Threshold changes, or deviations, however were treated as ordinal categorical variables because the preponderance (> 90 %) of the test-retest deviations fell within the range of -5 and 5 dB. To save time, initial models were prepared assuming an underlying continuum and final models utilized the multilevel ordinal logistic regression procedure implemented in the *gllamm* (Rabe-Hesketh & Skrondal 2012) add-on to Stata. Predictors of a direction of change were assessed using multilevel ordinal logistic regression. Predictors of the probability of a change in absolute value greater than 5 dB were assessed using multilevel binary logistic regression.

Finally, we reasoned that it was possible that the 1/3 OB stimuli might reduce the amount of observed change in threshold across neighboring frequencies more than either the ERB or pure tone stimuli. Thus, we derived slopes (dB/octave) for each frequency relative to the next lower frequency to identify whether stimulus bandwidth had an effect on audiogram slope. In these cases, slope was treated as a continuous variable and the data were analyzed using a multilevel regression model.

## Results

The majority of the participants had good hearing. The 75<sup>th</sup> percentile for pure tone thresholds (Table 2) was 15 dB HL or less through 3 kHz, and then declined to 25 dB HL at 6 and 8 kHz, but thresholds from one participant typically exceeded 80 dB HL at 8 kHz (Figure 3). The overall distributions of pure tone thresholds were similar across transducers and stimulus frequency. However, the marginal means at 0.5 kHz differed by stimulus in the final inferential model (described below). The marginal mean threshold for 1/3 OB noises at 0.5 kHz was somewhat lower than with the ERB and tone stimuli (Figure 4), and the interquartile range for 1/3 OB stimuli was also greater at 0.5 kHz.

The differences in mean threshold across transducers and stimuli provide a straightforward transformation of thresholds from a given combination of earphone and stimulus. These values (Table 3), rounded to the nearest 0.5 dB, can be summed with the observed threshold with any of the stimuli presented using the HDA200 earphone to achieve the best estimate of an equivalent pure tone threshold likely to have been obtained with the TDH-39P. In nearly all cases, these difference were within 2.5 dB and would therefore match the TDH-39P

thresholds obtained using a 5-dB audiometric step. The 1/3 OB noise at 0.5 kHz and the ERB noise at 4 kHz were exceptions and thresholds obtained with these stimuli would require adjustments by −5 and 5 dB, respectively.

**Deviations from baseline threshold**—Deviations from baseline threshold had medians of 0 dB in all cases. The interquartile ranges (Figure 5) were 5 dB or less except for pure tone thresholds at 6 and 8 kHz obtained using the TDH-39P transducers (Interquartile range = 10). Standard deviations of test-retest differences (Table 4) ranged between 7.1 dB (TDH-39P with tones at 8 kHz) and 3.3 dB (HDA200 with 1/3 OB at 1 kHz). Above 3 kHz, standard deviations tended to be lower via stimuli delivered from the HDA200 earphones, and an additional reduction in the standard deviation was observed for 1/3 OB signals at 6 and 8 kHz.

The standard deviations of test-retest differences were approximately 1.6 dB (range: 1.2 to 2.7 dB) greater than the standard deviations of the thresholds across repeated measurements at the same frequency for the same participant (Table 4). This was expected on the grounds that the expected variance of a difference is determined by the summed variances and the sum of the covariances of the variables contributing to the difference (see Nunnally & Bernstein, 1994). Although the standard deviations of the differences were greater, the rank-orders of the standard deviations of the differences were consistent with the rank-orders of the standard deviations of the thresholds ( $r = 0.97$ ), which was also expected because both rank-orderings were derived from the same underlying data.

Critical differences represent the dB difference that must be exceeded before one can conclude that a change has occurred with a stated level of confidence. Critical differences are specified via percentile points on the test-retest difference distribution. The 80 % critical differences were determined using the 10<sup>th</sup> and 90<sup>th</sup> percentiles of the difference distribution (Table 5). The 90 % critical differences (i.e., the 5<sup>th</sup> and 95<sup>th</sup> percentiles) were −5 to +5 dB for all frequencies, stimuli, and earphones through 2 kHz. Above 3 kHz, the TDH-39P earphones with tone signals had 90 % critical differences of [−10, +10], while all signals delivered via the HDA200 earphones tended to be [−5, +5]. These results and the inferential analyses that follow suggest that high frequency thresholds obtained with HDA200 earphones were more reliable than pure tone thresholds obtained with the TDH-39P.

## Inferential Results

**Stimulus and transducer effects on threshold**—The analyzed data consisted of 8197 observations of threshold across 294 tests of four conditions, 147 lab visits, and 49 participants. There were 35 thresholds missing from these data due to premature cessation of the protocol (one participant, 21 observations) and failure to conduct one test in the sequence according to the study protocol (two participants, 7 observations each). The multilevel model for threshold consisting of fixed factors for transducer, stimulus, the interaction between stimulus and frequency, frequency, and age in decades revealed significant differences across the fixed factors ( $\chi^2_{25} = 1138$ ;  $p < 0.00005$ ). The random factors of test and visit were unimportant, having an upper 95 % confidence interval boundary of less than 0.0045 dB, which suggests that tests within a visit and visits within

participants did not bear a systematic relationship with thresholds. A substantial random effect of participant having a standard deviation of 9.39 dB (95 % confidence interval [6.4, 13.7]) was observed, which illustrated the importance of accounting for the correlations among observations obtained from the same participant in the statistical model.

The main effects of frequency and age on thresholds were expected, and these factors were included only for statistical control. Stimulus, the interaction between frequency and stimulus, frequency, and age in decades were significant correlates of threshold in the multilevel mixed effects model (Table 6). No substantial effect of transducer was observed (coefficient = 0.12;  $p = 0.558$ ). The 95 % confidence interval for a mean difference between pure tone thresholds obtained with the TDH-39P and HDA200 ranges between  $-0.28$  and  $0.51$  dB, neither of which were statistically significant or practically important.

A significant main effect for stimulus was observed, but the main effect of stimulus is not interpretable without consideration of the interaction between stimulus and frequency. The interaction between frequency and stimulus indicates that differences in stimuli played more of a role at some frequencies than others. Homogeneous subsets (within-subset  $p > .05$ ) of threshold groups were derived and revealed some significant threshold differences across stimuli within frequencies. These analyses (Table 6) were conducted only on the HDA200 data to avoid any biasing effect of comparing thresholds obtained with HDA200 noise bands to the mean pure tone thresholds obtained using both earphone models, and corresponding marginal mean values are represented in Figure 4. At 0.5 and 1 kHz, 1/3 OB thresholds were significantly higher (i.e., apparently worse) than thresholds obtained with either pure tones or ERBs, which were not significantly different from each other. No significant threshold differences across stimuli were observed at 2, 3, or 4 kHz. At 6 and 8 kHz, pure tone thresholds were significantly greater (i.e., apparently worse) than thresholds obtained with either ERBs or 1/3 OB stimuli, which were not significantly different from each other.

**Stimulus effects on threshold deviations**—The associations between threshold deviations (calculated as the differences between thresholds obtained during the first test on the first visit and subsequent observations) and stimulus and transducer characteristics were assessed multiple ways. First, signed differences (i.e., observed minus expected differences, preserving sign) were used as the dependent variable in order to identify whether these factors were associated with a tendency toward increases or decreases in thresholds across repeated observations. These analyses were conducted using a multilevel ordinal logistic regression model. Second, a binary variable derived from unsigned (i.e., absolute) deviations was used to identify factors related to deviations in either direction. The binary variable was coded so that all absolute deviations less than or equal to 5 dB were assigned one category while deviations greater than 5 dB took the other. Results obtained from the binary variable could help identify whether one type of earphone or stimulus will result in more dependable observations.

Analyses of signed deviations indicated that the only stimulus frequency had a significant influence on signed deviations. There was no significant effect of transducer (Odds Ratio = 1.10; 95 % CI [0.97, 1.26];  $p = .138$ ). Follow-up comparisons of stimuli revealed that the central tendency of signed deviations were not influenced significantly by the stimulus ( $\chi^2_1$

= 0.23;  $p = .633$ ). Frequency was also associated with signed deviations in the analyses of the 8 kHz reliability study (Flamme et al., 2014), and since that was the larger parent project for the current study, readers may refer to those analyses for further explanation of those relationships.

Analyses of the binary variable identifying deviations greater than 5 dB revealed a main effect for transducer and an interaction between stimulus and frequency (Table 7). As suggested in Figure 5 above, absolute deviations greater than 5 dB were more frequent with the TDH-39P than with the HDA200 earphones (Odds ratio: 2.8; 95 % CI [2.19, 3.65];  $z = 7.95$ ;  $p < .0005$ ). In pairwise comparisons of combinations of frequency and stimulus, threshold deviations greater than 5 dB at 2 kHz were less likely with pure tone stimuli than with either ERB or 1/3 OB stimuli, which were not significantly different from one another. At 6 and 8 kHz, threshold deviations greater than 5 dB were less likely with 1/3 OB stimuli than with pure tone or ERB stimuli, which were not significantly different from one another.

**Effect of bandwidth on audiogram slope**—The potential effect of noise bands on audiogram slope was explored using a multilevel regression model in which the observed threshold, nominal frequency, stimulus, and the interaction between frequency and stimulus were used to predict audiogram slope. Slopes were defined as unsigned (i.e., absolute) dB/octave for frequencies of 1 kHz and above. The slope value was calculated as the dB difference between the selected frequency and the next lower frequency.

One cannot have a notch without having reduced hearing sensitivity at the notch frequency, so this issue is only relevant for tests showing substantial differences in threshold across frequency. Great threshold differences across neighboring frequencies were not common in these data, and we wished to reduce the extent to which the outcomes were dominated by essentially flat slopes. Analyses for this question were limited to threshold differences of 15 dB or more at adjacent frequencies, which corresponds to a slope greater than 25 dB per octave, which corresponds to changes greater than 10.4 dB between neighboring frequencies in cases the lower frequency is an inter-octave frequency (e.g., 3 kHz) and 14.6 dB in cases where the lower frequency is an octave frequency (e.g., 4 kHz). This reduced the data set for this analysis to 496 observations obtained from 172 tests from 38 people.

Audiometric slopes were related to the threshold, stimulus, and frequency (Table 8). The main effect of stimulus (Figure 6) was consistent with the hypothesis of reduced audiometric slope as a function of increased bandwidth beyond the ERB. The contrast between slopes observed with pure tone stimuli versus 1/3 OB stimuli was significant ( $p = 0.03$ ) when evaluated using conventional standard errors, but failed to reach statistical significance ( $p = 0.06$ ) when evaluated using robust standard errors. We report the effect as significant because the regression model that included stimulus as a factor provided a significantly better fit to the data (change in model  $\chi^2_2 = 3287$ ;  $p < .005$ ) than the regression model without stimulus as a factor.

## Discussion

The purposes of this study were to (1) determine the transformation from thresholds obtained with the HDA200 circumaural earphone into equivalent thresholds obtained using the TDH-39P supra-aural earphone, (2) compare the reliability of pure tone thresholds obtained with TDH-39P earphones and thresholds obtained with HDA200 earphones using pure tone, ERB, and 1/3 OB noise stimuli, and (3) explore the impact of signal bandwidth on the audiometric slope observed between neighboring frequencies. These results suggested that minimal transformation is needed to transfer ERB or 1/3 OB thresholds obtained with HDA200 earphones into the equivalent values that likely would have been obtained using TDH-39P earphones (Table 3). This result is similar to prior work (e.g., Cox & McDaniel, 1986). The reliability of thresholds obtained with the HDA200 earphones was superior to that obtained using TDH-39P earphones, especially in the high frequencies. Minimal differences were observed across stimuli in this study, but the observed differences suggest that the ERB noises produce pure tone thresholds with generally comparable central tendencies and reliability to those obtained with pure tones. The 1/3 OB noises might be somewhat more reliable than pure tones and ERB signals, but this added reliability comes at the cost of frequency resolution. The 1/3 OB noises tended to yield slightly shallower audiometric slopes on audiograms containing significant slopes, which would result in the reduction of notch depth and could reduce the detectability of focal damage by exciting auditory channels adjacent to the nominal stimulus frequency. This issue is perhaps of minimal importance in the assessment of hearing protector effectiveness, but it could result in delayed identification of new cases of hearing impairment during the audiometric monitoring phase of a hearing conservation program. The use of ERB noises would be a good compromise that might allow the testing of both earplug attenuation and hearing thresholds within a combined test protocol.

Perhaps the most striking result of this study is that the use of the circumaural HDA200 earphone provided 90 % critical differences (Table 5) that were always equal to or better than those obtained with the TDH-39P, with the greatest reliability improvement in the high frequencies. This result was present in the threshold data despite slightly poorer calibration consistency with the HDA200 earphones (Table 1). It is possible that the increased variability in daily calibration values might be an artifact of mounting the HDA200 on the flat plate, specifically the continued compaction of the HDA200 earphone cushion against the flat plate over time. This possibility was explored (unpublished data) by measuring the output of the HDA200 earphones as a function of time on both the Type 43AA flat plate/ear simulator assembly and a manikin head (GRAS Type 43AC). Sound levels increased linearly as a function of logarithmic time on the flat plate - particularly at 3 and 4 kHz - but minimal change was observed on the manikin. The variability observed in daily calibration values could be related to slight differences in the time interval between the placement of the earphone on the plate and the measurement. Regardless of the reason for the calibration variability, the circumaural headphone has demonstrated improved reliability for assessing high frequency thresholds compared to the supraaural headphone. Since the Sennheiser HDA200 headphone is no longer commercially available the ANSI standard should be updated and identify headphones that have equivalent performance characteristics to the

HDA200 with regards to attenuation, frequency response, dynamic range and distortion. The selection of headphone should be carefully considered especially strong in the context of occupational audiometry because of the increased ambient noise attenuation available via the circumaural enclosure.

Although the HDA200 earphones are no longer manufactured, the results of this study indicate that earphones using a circumaural enclosure are capable of providing threshold data that have equal or better reliability than the TDH-39P. Test-retest differences are especially important in the context of audiometric monitoring and for field-testing of the attenuation of hearing protectors, and the small corrections required to achieve equivalent TDH-39P thresholds (Table 3) and the magnitudes of the test-retest differences show that it is feasible to combine audiometric monitoring and earplug fit-testing in field environments. Additional studies are needed to identify models of circumaural earphones for this purpose. These studies should establish correction factors relative to TDH-39P earphones, expected test-retest differences, and ambient noise attenuation values for each candidate earphone model.

While there is reason to suspect that reduced slopes and/or notch depth could be obtained with 1/3 OB noise bands, we cannot rule out the possibility that a similar effect could be noticed in some cases with the ERB stimuli used in this study. The ERB bandwidths used in this study were obtained for moderate-level stimuli. On the basis of expected changes in basilar membrane excitation as a function of level, narrower ERBs would be expected for lower-level stimuli and broader ERBs would be expected for higher-level stimuli. A future study to derive optimal bandwidths for either low- or mid-level ERBs should be considered, and the participants in such a study would ideally be selected to oversample people having notched and steeply-sloping audiograms and such a study could also benefit from threshold testing conducted at a higher resolution in the high frequency (2–8 kHz) region, where notches are most commonly present.

## Acknowledgments

This study was supported by CDC/NIOSH Contract number 254-2011-M-40487. We also acknowledge the contributions of Amanda Hessenauer and Devon VanGessel, who helped develop the parent study procedures that were modified for use in this study.

The data from this manuscript and the associated analyses were presented at the National Hearing Conservation Association meeting held in Las Vegas in March 2014. A detailed contractor report was provided to the National Institute for Occupational Safety and Health and is available upon request.

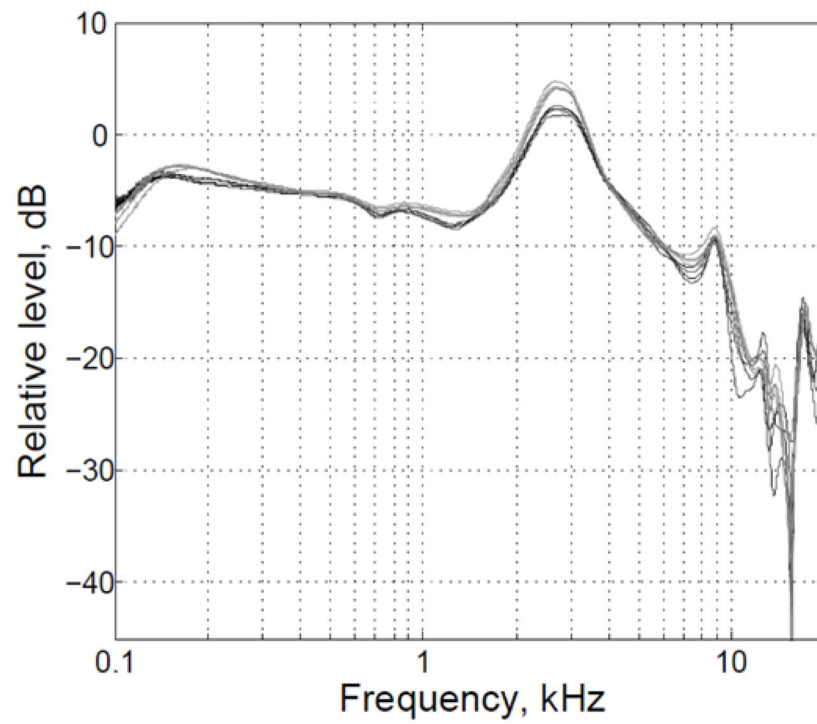
## Abbreviations

<b>ANSI</b>	American National Standards Institute
<b>ART</b>	Audiometric Research Tool
<b>CDC</b>	US Centers for Disease Control and Prevention
<b>CI</b>	Confidence interval
<b>dB</b>	decibel

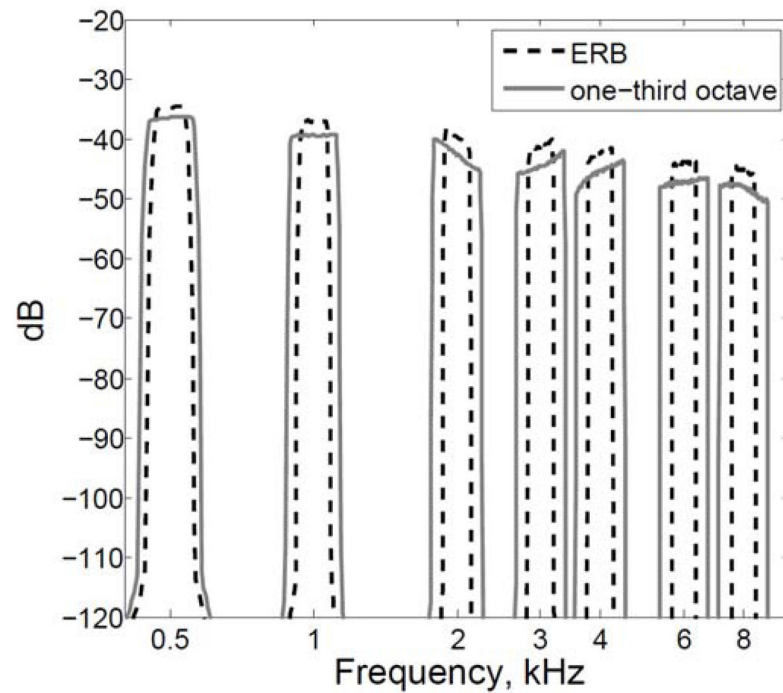
<b>ERB</b>	Equivalent rectangular band
<b>HDA200</b>	Sennheiser HDA200 circumaural earphone
<b>HL</b>	Hearing threshold level
<b>KEMAR</b>	Knowles Electronics Manikin for Acoustic Research
<b>kHz</b>	kilohertz
<b>NI</b>	National Instruments
<b>NIOSH</b>	National Institute for Occupational Safety and Health
<b>OB</b>	Octave band
<b>RETSPL</b>	Reference equivalent threshold SPL
<b>SD</b>	standard deviation
<b>SE</b>	standard error of the estimate
<b>SPL</b>	sound pressure level
<b>TDH-39P</b>	Telephonics Dynamic Headphone model 39 with plastic case

## References

- ANSI S3.1-1999 (R2008). Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms. New York, NY: American National Standards Institute, Inc;
- ANSI S3.6-2010. Specification for Audiometers. Melville, New York: Acoustical Society of America;
- Berger EH, Franks JR, Behar A, Casali JG, Dixon-Ernst C, Kieper RW, Merry CJ, Mozo BT, Nixon CW, Ohlin D, Royster JD, Royster LH. Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices, Part III. The validity of using subject-fit data. *J Acoust Soc Am*. 1998; 103:665–672. [PubMed: 9479749]
- Cox RM, McDaniel DM. Reference equivalent threshold levels for pure tones and 1/3-oct noise bands: Insert earphone and TDH-49 earphone. *J Acoust Soc Am*. 1986; 79:443–446. [PubMed: 3950196]
- Engdahl B, Tambs K, Hoffman HJ. Otoacoustic emissions, pure tone audiometry, and self-reported hearing. *Int J Audiol*. 2012; 52:74–82. [PubMed: 23216196]
- Flamme GA, Stephenson MR, Deiters KK, Hessenauer A, VanGessel DK, Geda K, Wyllys K, McGregor KD. Short-term variability of pure-tone thresholds obtained with TDH-39P earphones. *Int J Audiol*. 2014; 53(Suppl 2):S5–15. [PubMed: 24564693]
- Fletcher H, Wegel RL. The frequency-sensitivity of normal ears. *Phys Rev*. 1922; 19:553–565.
- Glasberg BR, Moore BCJ. Derivation of auditory filter shapes from notched-noise data. *Hear Res*. 1990; 47:103–138. [PubMed: 2228789]
- Huber, PJ. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. Berkeley: University of California Press; 1967. The behavior of maximum likelihood estimates under nonstandard conditions; p. 221-233.
- Murphy WJ. Comparing personal attenuation ratings for hearing protector fit-test systems. *CAOHC Update*. 2013; 25(3):6–8.
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. 3. New York: McGraw-Hill; 1994. p. 159-208.
- Rabe-Hesketh, S.; Skrondal, A. *Multilevel and Longitudinal Modeling using Stata*. 3. Vol. 2. College Station, TX: Stata Press; 2012.
- Unoki M, Irino T, Glasberg B, Moore BC, Patterson RD. Comparison of the roex and gammachirp filters as representations of the auditory filter. *J Acoust Soc Am*. 2006; 120:1474–1492. [PubMed: 17004470]

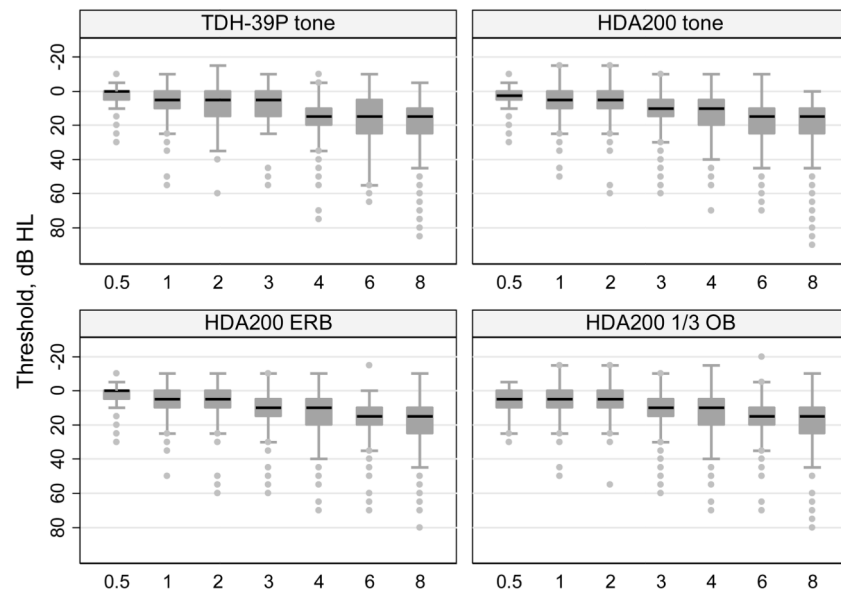


**Figure 1.**  
Frequency responses of the HDA200 earphones on KEMAR.



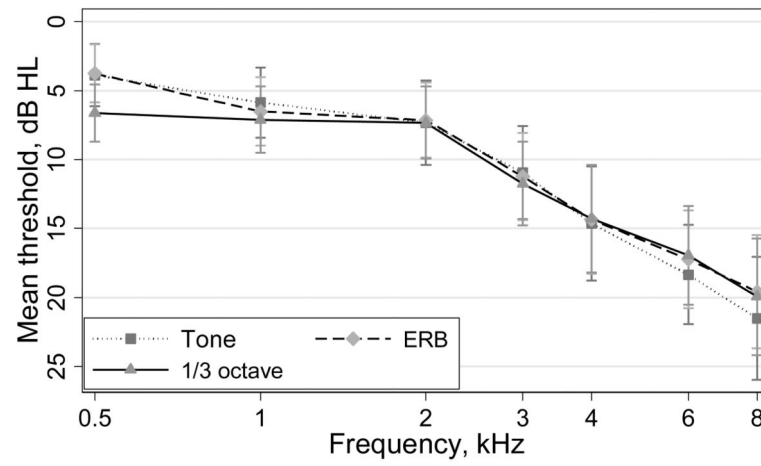
**Figure 2.**

Spectra of noise stimuli stored on .wav files. Note that the passbands of the noise signals are not flat due to the inverse filtering to adjust for the HDA200 frequency response at KEMAR's eardrum. The spectrum within the passband is lower for 1/3 OB stimuli because the stimuli were presented at an equal overall level rather than an equal spectrum level.

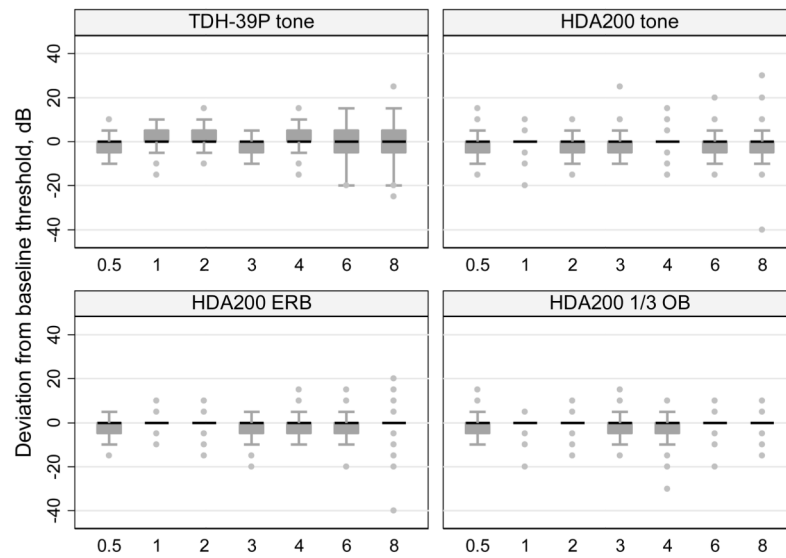


**Figure 3.**

Boxplot of threshold distributions as a function of frequency (kHz) by transducer and stimulus. Gray boxes represent interquartile (i.e., 25<sup>th</sup> to 75<sup>th</sup> percentile) ranges. Black lines represent medians. Bars represent the upper and lower adjacent values and circles represent observations outside the range of the upper and lower adjacent values.

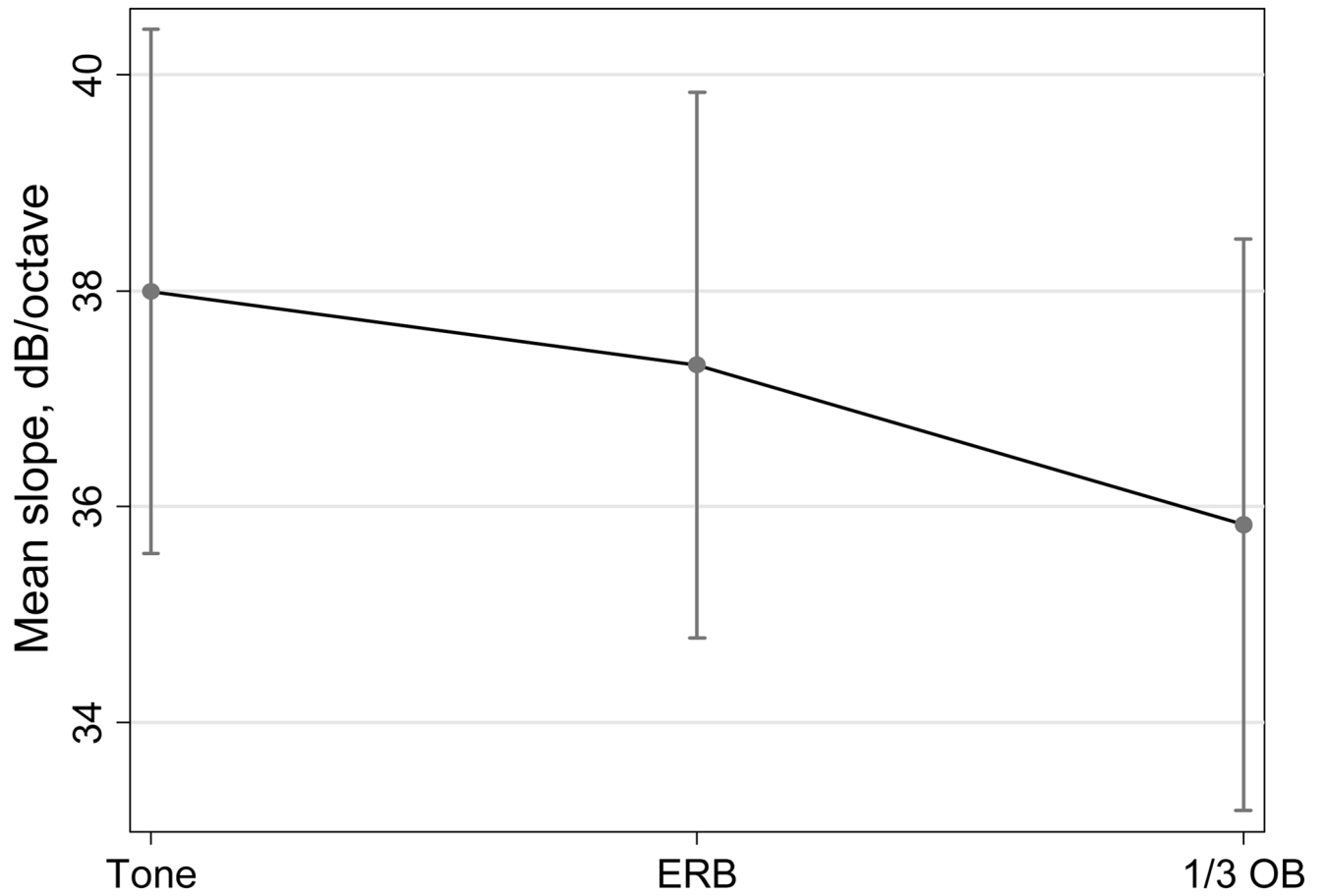


**Figure 4.** Mean thresholds by frequency and stimulus for stimuli delivered via the HDA200 earphone. Error bars represent the 95 % confidence intervals for the means.



**Figure 5.**

Threshold deviations as a function of transducer, stimulus, and frequency. Gray boxes represent interquartile (i.e., 25<sup>th</sup> to 75<sup>th</sup> percentile) ranges. Black lines represent medians. Bars represent the upper and lower adjacent values and circles represent observations outside the range of the upper and lower adjacent values. Interquartile ranges and adjacent values are not visible in cases where the interquartile range is compressed into a single observed level (e.g., a 0 dB deviation).



**Figure 6.**  
Mean slope, in dB per octave, between neighboring frequencies as a function of stimulus.  
Error bars represent the 95 % confidence interval for the mean, calculated using robust standard errors. Slightly shallower slopes were observed with the 1/3 OB stimulus.

Table 1

Standard deviations of daily calibration values, by transducer.

	kHz	0.5	1	2	3	4	6	8
TDH-39P	left	0.1	0.2	0.1	0.1	0.1	0.3	0.4
	right	0.1	0.1	0.1	0.1	0.1	0.2	0.3
HDA200	left	0.3	0.4	0.4	0.8	0.7	0.4	0.4
	right	0.3	0.4	0.4	0.7	0.7	0.4	0.4

Table 2

Threshold percentiles, means and standard deviations across all 8197 threshold observations as a function of frequency.

Frequency, kHz	0.5	1	2	3	4	6	8
25 <sup>th</sup> percentile	0	0	0	5	5	10	10
50 <sup>th</sup> percentile	5	5	5	10	10	15	15
75 <sup>th</sup> percentile	5	10	10	15	20	25	25
Mean	4.4	6.4	7.6	10.9	14.9	17.4	20.3
SD	8.0	9.6	11.5	12.6	15.4	13.9	16.9

**Table 3**  
Correction factors from HDA200 to TDH-39P tone thresholds, rounded to the nearest 0.5 dB.

<b>kHz</b>	<b>0.5</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>6</b>	<b>8</b>
Tones	-0.50	0.50	1.50	-1.00	2.00	-1.00	-1.50
ERB	0.00	0.50	2.00	-1.00	2.50	0.50	1.00
1/3 OB	-3.00	-0.50	1.50	-2.00	2.50	0.50	0.50

**Table 4**

Descriptive statistics for test-retest differences as a function of frequency, transducer, and stimulus. The standard deviation (SD) of differences was calculated relative to the baseline (i.e., first test on first visit) threshold. The standard deviation of thresholds was calculated as the mean of the participant-specific standard deviations across all observations.

kHz	Transducer	Stimulus	Percentile					Mean	SD of Differences	SD of Thresholds
			1	25	50	75	99			
0.5	TDH-39P	Tone	-10	-5	0	0	10	-1.2	4.0	2.5
		Tone	-10	-5	0	0	10	-0.7	3.7	2.2
	HDA200	ERB	-10	-5	0	0	5	-0.9	3.3	1.8
		1/3 OB	-5	-5	0	0	10	-0.8	3.5	2.3
1	TDH-39P	Tone	-15	0	0	5	10	0.4	4.0	2.5
		Tone	-10	0	0	0	10	-0.4	3.9	2.5
	HDA200	ERB	-5	0	0	0	10	-0.4	3.4	2.1
		1/3 OB	-10	0	0	0	5	-0.9	3.3	2.1
2	TDH-39P	Tone	-5	0	0	5	10	1.1	3.9	2.4
		Tone	-10	-5	0	0	5	-0.5	4.0	2.8
	HDA200	ERB	-10	0	0	0	5	-0.3	3.8	2.2
		1/3 OB	-10	0	0	0	10	-0.5	3.8	2.3
3	TDH-39P	Tone	-10	0	0	0	5	-2.0	3.9	2.2
		Tone	-10	0	0	5	10	-1.0	4.2	2.7
	HDA200	ERB	-15	-5	0	5	5	-0.9	4.2	2.6
		1/3 OB	-15	-5	0	5	10	-0.6	4.1	2.7
4	TDH-39P	Tone	-15	-5	0	5	15	1.4	5.5	3.2
		Tone	-15	-5	0	0	10	-0.5	4.1	2.7
	HDA200	ERB	-10	-5	0	0	10	-0.4	4.4	2.7
		1/3 OB	-15	0	0	0	5	-0.6	4.7	2.9
6	TDH-39P	Tone	-15	-5	0	5	10	-1.7	6.3	4.2
		Tone	-10	-5	0	0	10	-0.7	4.5	2.8
	HDA200	ERB	-10	-5	0	0	10	-0.6	4.5	2.8
		1/3 OB	-10	0	0	0	10	0.2	3.8	2.5

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

kHz	Transducer	Stimulus	Percentile					Mean	SD of Differences	SD of Thresholds
			1	25	50	75	99			
8	TDH-39P	Tone	-20	-5	0	5	15	-1.2	7.1	4.4
		Tone	-10	-5	0	0	20	-0.1	5.6	3.6
	HDA200	ERB	-15	0	0	0	10	-0.1	5.5	3.4
		1/3 OB	-10	0	0	0	10	-0.2	3.7	2.5

**Table 5**

80% and 90% critical differences for thresholds, by frequency, transducer, and stimulus.

kHz	Transducer	Stimulus	80 % Critical Difference		90 % Critical Difference	
			Low	High	Low	High
0.5	TDH-39P	Tone	-5	5	-5	5
		Tone	-5	5	-5	5
	HDA200	ERB	-5	5	-5	5
		1/3 OB	-5	5	-5	5
1	TDH-39P	Tone	-5	5	-5	5
		Tone	-5	5	-5	5
	HDA200	ERB	-5	5	-5	5
		1/3 OB	-5	0	-5	5
2	TDH-39P	Tone	-5	5	-5	5
		Tone	-5	5	-5	5
	HDA200	ERB	-5	5	-5	5
		1/3 OB	-5	5	-5	5
3	TDH-39P	Tone	-5	0	-10	5
		Tone	-5	5	-10	5
	HDA200	ERB	-5	5	-10	5
		1/3 OB	-5	5	-5	5
4	TDH-39P	Tone	-5	5	-10	10
		Tone	-5	5	-5	5
	HDA200	ERB	-5	5	-5	5
		1/3 OB	-5	5	-10	5
6	TDH-39P	Tone	-10	5	-10	10
		Tone	-5	5	-5	5
	HDA200	ERB	-5	5	-5	5
		1/3 OB	-5	5	-5	5
8	TDH-39P	Tone	-10	5	-10	10
	HDA200	Tone	-5	5	-5	5

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

kHz	Transducer	Stimulus	80 % Critical Difference		90 % Critical Difference	
			Low	High	Low	High
		ERB	-5	5	-10	5
		1/3 OB	-5	5	-5	5

**Table 6**

Multilevel regression coefficients and confidence intervals for the association between observed threshold and transducer, stimulus, frequency, and age. The reference group (i.e., the condition for which only the intercept coefficient applies) is 20 to 29 year olds tested using a 0.5 kHz pure tone presented via HDA200 earphones. Model 1 includes all thresholds, model 2 includes only thresholds obtained using HDA200 earphones. Robust standard errors were used when calculating confidence intervals. Statistical significance ( $p < 0.05$ ) of coefficients is present when the 95 % CI does not include zero.

	Model 1	95 % Confidence Interval		Model 2	95 % Confidence Interval	
	Coefficient	Low	High	Coefficient	Low	High
Transducer						
TDH-39P	0.12	-0.28	0.51	--	--	--
Stimulus						
ERB	0.11	-0.45	0.68	-0.12	-0.73	0.49
1/3 OB	2.99	2.47	3.50	2.76	2.29	3.22
Stimulus*Frequency						
ERB, 1 kHz	0.24	-0.59	1.06	0.77	-0.13	1.66
ERB, 2 kHz	-1.00	-1.91	-0.09	-0.06	-1.01	0.88
ERB, 3 kHz	0.74	-0.16	1.64	0.39	-0.61	1.39
ERB, 4 kHz	-1.34	-2.40	-0.30	-0.18	-1.17	0.80
ERB, 6 kHz	-0.74	-1.63	0.16	-1.00	-1.96	-0.04
ERB, 8 kHz	-1.39	-2.67	-0.12	-1.80	-3.11	-0.49
1/3 OB, 1 kHz	-2.04	-2.72	-1.37	-1.51	-2.23	-0.80
1/3 OB, 2 kHz	-3.69	-4.70	-2.70	-2.76	-3.72	-1.79
1/3 OB, 3 kHz	-1.60	-2.56	-0.65	-1.96	-2.93	-0.98
1/3 OB, 4 kHz	-4.28	-5.49	-3.07	-3.11	-4.22	-2.00
1/3 OB, 6 kHz	-3.89	-4.90	-2.88	-4.15	-5.10	-3.19
1/3 OB, 8 kHz	-3.91	-5.21	-2.61	-4.32	-5.53	-3.11
Frequency						
1 kHz	2.52	1.05	3.99	1.99	0.45	3.52
2 kHz	4.39	1.51	7.27	3.45	0.59	6.31
3 kHz	6.72	3.61	9.84	7.07	3.86	10.28
4 kHz	11.93	8.13	15.73	10.77	6.92	14.61

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Model 1	95 % Confidence Interval		Model 2	95 % Confidence Interval	
	Coefficient	Low	High	Coefficient	Low	High
6 kHz	14.21	11.00	17.43	14.47	11.02	17.92
8 kHz	17.23	13.31	21.14	17.64	13.52	21.75
Age, decade						
30–39	8.01	2.28	13.75	8.15	2.43	13.87
40–49	6.72	1.97	11.47	6.64	1.95	11.32
50–59	8.84	5.71	11.97	9.06	5.92	12.20
60–69	19.30	8.44	30.16	19.29	8.42	30.15
Intercept	–5.52	–7.76	–3.27	–5.34	–7.60	–3.08

**Table 7**

Effect of stimulus and frequency on the probability of absolute deviations greater than 5 dB. Odds ratios were obtained via multilevel logistic regression. Robust standard errors were used when calculating confidence intervals. Statistical significance ( $p < 0.05$ ) of coefficients is present when the 95 % CI does not include a value of 1.0.

	Odds Ratio	95 % Confidence Interval	
		Low	High
Transducer			
TDH-39P	2.83	2.19	3.65
Stimulus			
ERB	1.32	0.61	2.84
1/3 OB	0.93	0.41	2.12
Stimulus*Frequency			
ERB, 1 kHz	0.99	0.31	2.54
ERB, 2 kHz	1.73	0.62	4.84
ERB, 3 kHz	0.92	0.35	2.43
ERB, 4 kHz	1.00	0.39	2.56
ERB, 6 kHz	0.53	0.21	1.33
ERB, 8 kHz	0.78	0.32	1.92
1/3 OB, 1 kHz	1.00	0.33	3.07
1/3 OB, 2 kHz	2.44	0.84	7.13
1/3 OB, 3 kHz	1.12	0.40	3.14
1/3 OB, 4 kHz	1.18	0.43	3.19
1/3 OB, 6 kHz	0.33	0.12	0.92
1/3 OB, 8 kHz	0.34	0.12	0.96
Frequency			
1 kHz	1.13	0.65	1.97
2 kHz	0.17	0.47	1.50
3 kHz	1.98	0.85	3.33
4 kHz	2.51	1.51	4.16
6 kHz	5.26	3.25	8.52
8 kHz	4.64	2.85	7.53
Intercept	0.00538	0.00214	0.01353

Table 8

Multilevel linear regression analysis of slope differences as a function of stimulus, controlling for threshold and frequency. Conventional standard errors were calculated using asymptotic theory, robust standard errors were based on the sandwich estimator (Huber, 1967). Statistical significance ( $p < 0.05$ ) of coefficients is present when the 95 % CI does not include zero.

	Coefficient	Conventional SE		Robust SE	
		95 % low	95 % high	95 % low	95 % high
Stimulus					
ERB	-0.68	-2.50	1.14	-2.29	0.92
1/3 OB	-2.17	-4.15	-0.19	-4.46	0.12
Threshold, dB	0.24	0.17	0.31	0.14	0.33
Frequency					
3 kHz	7.67	-10.08	25.43	2.73	12.62
4 kHz	20.19	2.68	37.70	17.13	23.24
6 kHz	9.62	-8.07	27.31	-0.25	19.50
8 kHz	16.17	-1.50	33.83	12.31	20.02
Intercept	17.50	-0.41	35.40	12.47	22.53